UNITED STATES PATENT APPLICATION FOR:

STEREO-VISION BASED IMMINENT COLLISION DETECTION

INVENTORS:

PENG CHANG

AND

THEODORE ARMAND CAMUS

ATTORNEY DOCKET NUMBER: SAR 14948

CERTIFICATION OF MAILING UNDER 37 C.F.R. 1.10

I hereby certify that this New Application and the documents referred to as enclosed therein are being deposited with the United States Postal Service on 1/29/04, in an envelope marked as "Express Mail United States Postal Service", Mailing Label No. EU137712996US, addressed to: Assistant Commissioner for Patents, Box PATENT APPLICATION, Washington, D.C. 20231.

Kimberly Boyd
Signature

Kimberly Boyd
Name

1/29/04
Date of signature

MOSER, PATTERSON & SHERIDAN LLP
595 Shrewsbury Ave.
Shrewsbury, New Jersey 07702
(732) 530-9404

STEREO-VISION BASED IMMINENT COLLISION DETECTION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of United States provisional patent application number 60/484463, filed July 2, 2003, entitled, "Stereo Vision Based Algorithms for Automotive Imminent Collision Avoidance," by Chang et al., which is herein incorporated by reference.

[0002] This application is a continuation-in-part of pending United States Patent Application serial number 10/461,699, filed on 6/13/2003, entitled, "VEHICULAR VISION SYSTEM" (Attorney Docket Number SAR14885), by Camus et al. That Patent application is hereby incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

Field of the Invention

[0003] The present invention relates to vision systems, e.g., as deployed on a vehicle. In particular, this invention relates to detecting imminent collisions using stereo vision.

Description of the Related Art

[0004] Significant interest exists in the automotive industry for systems that detect imminent collisions in time to avoid that collision or to mitigate its damage. Collision avoidance systems typically must detect the presence of potential threats, determine their speed and trajectory, and assess their collision threat. Prior art collision avoidance systems have used radar to determine the range and closing speed of potential threats. However, affordable radar systems usually lack the required spatial resolution to reliably and accurately determine the size

and the location of potential threats.

[0005] Since stereo vision can provide the high spatial resolution required to identify potential threats, stereo vision has been proposed for use in collision detection and avoidance systems. For example, US Patent Application 10/461,699, filed on June 13, 2003 and entitled "VEHICULAR VISION SYSTEM," which is hereby incorporated by reference in its entirety, discloses detecting and classifying objects (potential threats) using disparity images, depth maps, and template matching. While the teachings of US Patent Application 10/461,699 are highly useful, its methods of detecting potential threats are not optimal in all applications.

[0006] Therefore, there is a need in the art for new techniques of using stereo vision for collision detection and avoidance.

SUMMARY OF THE INVENTION

[0007] In one embodiment the principles of the present invention provide for stereo vision-based collision detection.

[0008] In one embodiment, a stereo vision based collision avoidance systems that is in accord with the present invention includes stereo cameras that produce imagery that is processed to detect vehicles within a field of view. Such processing includes determining the size, speed and direction of potential threats and an assessment of the collision threat posed by the detected potential threats.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] So that the manner in which the above recited features of the present invention are attained and can be understood in detail, a more particular description of the invention, briefly summarized above, may be had by reference to the embodiments thereof which are illustrated in the appended drawings.

[0010] It is to be noted, however, that the appended drawings illustrate only

typical embodiments of this invention and are therefore not to be considered limiting of its scope, for the invention may admit to other equally effective embodiments.

[0011] Figure 1 depicts a schematic view of a vehicle having a stereo vision system that is in accord with the principles of the present invention;

[0012] Figure 2 illustrates electronic subsystems of the stereo vision system of Figure 1;

[0013] Figure 3 depicts producing a depth map;

[0014] Figure 4 depicts a flow chart of a potential threat detection and segmentation process used in the vision system of Figure 2,

[0015] Figure 5 depicts a top view of an imminent collision;

[0016] Figure 6 depicts a flow chart of a plane fitting and labeling process used in the potential threat detection and segmentation process of Figure 4;

[0017] Figure 7 depicts a flow chart of a velocity estimation process used in the vision system of Figure 3, and

[0018] Figure 8 depicts a flow chart for performing velocity estimation.

DETAILED DESCRIPTION

[0019] A primary requirement of a collision avoidance system is the detection of actual collision threats to a platform, e.g., a host vehicle. Once an imminent collision is detected the host vehicle (platform) may take action either to avoid the collision and/or to mitigate the damage caused by the collision. Information regarding the size, location, and motion of a potential threat is useful in determining if a specific measure that could be taken is appropriate under the given conditions.

[0020] A collision detection system that is in accord with the principles of the present invention estimates the location and motion of potential threats, determines various properties of those threats, such as size, height, and width, and classifies the potential threats to identify imminent collisions based upon the previously estimated location, motion, and properties. Since collision detection often involves vehicles traveling at high speed, a collision detection system that is in accord with the principles of the present invention incorporates efficiently executed algorithms that are sufficiently robust to accommodate a wide range of potential threats, lighting conditions, and other circumstances.

[0021] Figure 1 depicts a schematic diagram of a host vehicle 100 having a collision detection system102 that is in accord with the principles of the present invention. That system detects potential threats within a scene 104 that is proximate the vehicle 100. That scene may include non-threatening objects such as a pedestrian 103 as well as potential threats, shown in Figure 1 as a vehicle 110. While Figure 1 shows the scene104 in front of the host vehicle 100, other collision detection systems may image scenes that are behind or to the side of the host vehicle 100. The collision detection system 102 uses a stereo vision imaging device 106 that is coupled to an image processor 108. The stereo vision imaging device 106 has a field of view that includes the pedestrian 103 and the vehicle 110.

[0022] Figure 2 depicts a block diagram of hardware used to implement the collision detection system 102. The stereo vision imaging device 106 comprises a pair of cameras 200 and 202 that operate in the visible wavelengths. The cameras have a known relation to one another such that they can produce a stereo image of the scene 104 from which information can be derived. The image processor 108 comprises an image preprocessor 206, a central processing unit (CPU) 210, support circuits 208, and memory 217. The image preprocessor 206 generally comprises circuitry for capturing, digitizing and processing the stereo imagery from the sensor array 106. The image

preprocessor may be a single chip video processor such as the processor manufactured under the model Acadia I™ by Pyramid Vision Technologies of Princeton, New Jersey.

[0023] The processed images from the image preprocessor 206 are coupled to the CPU 210. The CPU 210 may comprise any one of a number of presently available high speed microcontrollers or microprocessors. The CPU 210 is supported by support circuits 208 that are generally well known in the art. These circuits include cache, power supplies, clock circuits, input-output circuitry, and the like. The memory 217 is also coupled to the CPU 210. The memory 217 stores certain software routines that are executed by the CPU 210 and by the image preprocessor 206 to facilitate the operation of the invention. The memory also stores certain databases 214 of information that are used by the invention, and image processing software 216 that is used to process the imagery from the sensor array 106. Although the invention is described in the context of a series of method steps, the method may be performed in hardware, software, or some combination of hardware and software.

[0024] Figure 3 is a block diagram of functional modules that determine if a collision is imminent. The cameras 200 and 202 provide stereo imagery for a stereo image preprocessor 300. The stereo image preprocessor 300 calibrates the cameras 200 and 202, captures and digitizes stereo imagery, warps the images into alignment, and performs a pyramid wavelet decomposition to create multi-resolution disparity images. Camera calibration is important as it provides a reference point and a reference direction from which all distances and angles are determined. In particular, the separation of the cameras is important since the disparity images contain the point-wise motion from the left image to the right image. The greater the computed disparity of a potential threat, the closer that threat is to the cameras 200 and 202. After preprocessing, a depth map generator 302 produces a depth map. The depth map contains data representative of the image points, where each point represents a specific

distance from the cameras 200 and 202 to a point within the scene 104. The depth map is used by a collision detector 304 that detects whether a collision is imminent using the processes that are described below.

[0025] Figure 4 depicts a flow diagram of the operation of the collision detection system 102. At steps 302 and 304 the stereo cameras 200 and 202 provide left and right image stream inputs that are processed at step 306 to form the stereo depth map using the method described with reference to Figure 3. With the stereo depth data available, at step 308 a threat detection and segmentation algorithm detects potential threats in the stereo depth data, and thus in the scene 104 (see Figure 1). The threat detection and segmentation step 308 returns "bounding boxes" of potential threats in the stereo depth data. The threat detection and segmentation algorithm used in step 308 is described in more detail subsequently.

[0026] Once bounding boxes are obtained, the properties of the potential threats can be obtained from the stereo depth data. At step 310 the size and height of the potential threats are determined; at step 312 the relative position of the potential threats are determined; and at steps 314 and 316 a velocity estimation algorithm is performed that provides velocity estimates for the potential threats. The details of determining those properties are described subsequently.

[0027] All of the properties determined in steps 310, 312, and 314-316 are estimates that are derived from the stereo depth data, which includes image noise. To reduce the impact of that noise, those property estimates are time filtered. More specifically, at step 318 the position and velocity measurements are filtered using Kalman filters, while at step 320 a low-pass filter filters noise from the other estimates. More details of filtering are provided subsequently. After low pass filtering, at step 322 the low pass filtered estimates are threshold detected. Threshold detection removes small and large objects from the potential threat list.

[0028] Once filtered size, position, and velocity estimates are known, at step 324

the collision avoidance system 102 performs a trajectory analysis and a collision prediction of the potential threats. That analysis, combined with the threshold determination from step 322, is used at step 326 to make a final decision as to whether an imminent collision with a potential threat is likely.

[0029] Figure 5 depicts a top view of a collision scenario in which a host vehicle 100 has identified a potential threat 110 as an imminent collision threat. A closing path 402 is represented by the line C-D, where C and D respectively represent the collision contact points of the potential threat 110 and the host vehicle 100. Points A and B represent the edges of the potential threat 110, which are determined in steps 310 and 320. The position of C can be computed once the positions of point A, B, D and the lateral and longitudinal velocities are known.

[0030] Turning back to step 308, threat detection and segmentation, that step is performed using a process (depicted in Figure 6) that processes the stereo depth map computed in step 306. Threat detection and segmentation is based on the principle that each pixel in the stereo depth map has an associated 3D position from which objects can be abstracted. However, the stereo depth map is often noisy and sparse. The flowchart depicted in Figure 6 starts with the stereo depth map data obtained at step 306. At step 602 that data is tessellated into a grid of patches. At step 604, for each patch a plane is fitted (in a manner that is subsequently explained) through data points within the specific patch, and then each patch is classified into predefined types. Those predefined types are based on the 3D positions of each patch and on the normal vector of the fitted plane. Broadly, the predefined types are of three general classes, those that are likely to represent a potential threat; those that possibly may represent a potential threat; and those that are unlikely to represent a potential threat. Step 604 is explained in more detail subsequently.

[0031] Still referring to Figure 6, at step 606 a grouping process groups the patches together based on their classifications. To reduce the effect of confusion

patches (a classification type that is described subsequently), the grouping process performs two groupings. In the first grouping the patches classified as car sides and car tops (see below) are grouped together. Those patches cover the potential threats. After the first grouping a second grouping adds confusion patches (additional patches usually at the boundary of the threat car) to the potential threats, if any exist. After grouping, at step 608 the classified patches are clustered to form bounding boxes. It should be noted that the second grouping of step 606 improves the clustering of the bounding boxes at potential threat boundaries.

[0032] Figure 7 illustrates the step 604 of plane fitting and patch classifying in more detail. Plane fitting is used to generate plane normals, which are important in the classification of the patches. At step 702 a patch is selected. Then, to mitigate problems caused by data insufficiency within the stereo data, at step 704 the patch is moved around locally to find the region of maximum stereo density near the original patch location. This reduces the effect of holes in the stereo data that cause problems such as increased errors when plane fitting. Holes, which represent pixels that do not have valid 3D position estimates, are caused by specularities, lack of texture, or other factors in the stereo image data. The 3D positions of the pixels can also contain noise and outliers, sometimes severe, which can also cause problems. Readily identifiable noise and outliers can also be removed from the stereo data. Then, at step 706 a determination is made as to whether the patch is dense enough to be used. If not, at step 708 a patch without sufficient density is discarded. Thus, not all patches are used in the collision avoidance system 102.

[0033] Still referring to Figure 7, at step 710, for each patch that is retained a subset of the stereo image data points for that patch is used for plane fitting and patch normal determination. For example, only pixels having depth values in the middle 80% of the overall range can be used. This eliminates possible outliers in the stereo data from skewing the results. Plane fitting starts by removing each

patch's distance offset from the stereo data. This forces the resulting patch plane to be such that the 3D position (x, y, z) of any point in the plane satisfies the equation $ax + by + cz = 0$, which is the desired plane equation having an origin at the patch center. Then, a plane is fit through the selected subset 3D points of each patch to form the desired patch plane. The resulting patch plane is such that for all points:

[0034] $Ax = 0$

[0035] where $x = (a, b, c)$ is the plane normal, and A is an N by 3 matrix with the 3-D coordinates with respect to the patch centroid, (x,y,z), for each point at every row. A least square solution of $Ax = 0$ provides the patch's (surface) normal vector. A computationally efficient way to calculate the surface normal vector is to calculate the third Eigen-vector of the matrix A^TA, by applying a singular valued decomposition (SVD) to the matrix A^TA. Fast SVD algorithms exist for positive semi-definite matrixes, which is the case for the matrix of interest.

[0036] Once the plane normal is available, at step 712 a decision is made as to whether to use the patch in collision detection. That decision is based on the classification of the patch, with the patch being classified as one of the following types:

[0037] a negative patch, if the patch has a negative height;

[0038] a ground patch, if the patch height is both below a threshold and has a vertical normal;

[0039] a faraway patch, if the patch distance is outside the scope of interest

[0040] a high patch, if the patch height is outside the scope of interest

[0041] a boundary patch, if the height is close to ground but has a non-vertical normal, or if the height is above the threshold but has a vertical normal;

[0042] a car side patch, if the height is above the threshold and has a non-vertical normal; or

[0043] a car top patch, if the height is above the threshold and with an almost vertical normal.

[0044] Patch classification is based on the orientation of the patch (as determined by its plane normal), on its height constraint, and on it position. Classifying using multiple criteria helps mitigate the impact of noise in the stereo image data. The exact thresholds to use when classifying depend on the calibration parameters of the cameras 200 and 202 and on the potential threats in the scene 104. In most cases the patches that are classified as car sides or car tops are, in fact, from a potential threat. Thus, the car side and car top classifications represent a general class of being from a potential threat. The confusion patches are often boundary patches which contain mixed parts of ground and a potential threat. Thus, confusion patches represent a general class that may represent a potential threat. If the patch is not classified as a car side, car top, or confusion patch the patch is unlikely to be from a potential threat and are thus discarded in step 714. However, if the patch is a car side, car top, or confusion patch, at step 716 the patch is marked as being part of a potential threat. Finally, after step 716, 714, or step 708, at step 718 a decision is made as to whether there are any more patches to classify. If yes, step 604 returns to select another patch at step 702. Otherwise, the method 300 proceeds to step 606.

[0045] The height, size and locations of potential threats can be measured directly from the bounding box. In particular, the left and right bounds of potential threats are determined over time to enable better estimates.

[0046] Figure 8 illustrates the methods of steps 314 and 316, velocity estimation, in more detail. After the bounding boxes in step 308 has been found the velocity of the potential threats can be estimated. First, at step 802 images of the scene 104 are obtained in different time frames. This is also shown by line 313 of Figure

4. Then, at step 804 2D correspondences for selected potential threats are determined using "feature matching" in the different time frames to establish 2D correspondences across time. Reference, D. Nister, "Five point structure from motion algorithm' Sarnoff Corporation, Invention Disclosure 14831. From the 2D correspondences alone the angular velocity of the potential threats can be determined.

[0047] Once the 2D correspondences are available, at step 806 the 3D correspondences for the same feature set can be found relatively easily using the depth changes for the same set of features in different frames. This produces two sets of data point sets, P_i and Q_i, wherein i=1...N and such that:

[0048] $Q_i = R\ P_i + T + V_i$,

[0049] where N is the number of data points, R is a rotation matrix, T is a 3D translation matrix, and V_i is noise.

[0050] Given two corresponding data point sets, at step 808 a 3D velocity estimate is obtained. Standard methods exist to solve for optimal rotation and translation motion if N is greater than two, see, for example, K. Arun, T. Huang, and S. Blostein, "Least-square Fitting of Two 3D Point Sets," IEEE Trans. Pattern Anal. Machine Intel., vol. 9, no. 5, pp. 698 (1987). For automotive collision avoidance it is often beneficial to assume pure translation motion. Then only translation motion needs to be estimated. However, a straight forward implementation of the method taught by K. Arun, T. Huang, and S. Blostein, leads to somewhat inferior results due to the existence of severe noise and outliers in the 3D correspondence data. Thus, step 808 uses a more robust method of estimating velocity based on Random Sample Consensus (RANSAC). The general algorithm is:

[0051] 1. Select k points from the 3D correspondence data sets:

[0052] 2. Solve for T (and optionally, R):

[0053] 3. Find how many points (out of N) fit within a tolerance, call it M.

[0054] 4. If M/N is large enough, accept the result and exit; otherwise

[0055] 5. Repeat 1to 4 L times or until M/N is large enough;

[0056] 6. Fail

[0057] It is possible to directly derive the 3D correspondences from the depth images using algorithms such as ICP (Iterative Closest Points). But, directly deriving 3D correspondences is computationally inefficient and is subject to noise in the stereo image data. The 3D correspondences provide the closing velocities of the potential threats relative to the host vehicle 100.

[0058] As noted above, to reduce the noise in the stereo data filtering is applied to all measurements. Measurements of constant quantities, such as potential threat size and height, are filtered using standard low pass filtering. Varying parameters, such as position and velocity measurements, are filtered using a Kalman filter. A system model is required by Kalman filter and a constant velocity model is used with an acceleration modeled as Gaussian white noise. The system motion equation may be written as:

$$\begin{bmatrix} X_l \\ Z_l \\ X_r \\ Z_r \\ V_x \\ V_z \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_l \\ Z_l \\ X_r \\ Z_r \\ V_x \\ V_z \end{bmatrix}_k + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ n_x \\ n_z \end{bmatrix}_k$$

[0059] All of the variables are directly measurable (except as explained below), therefore the observation equation is simply the variables themselves plus the measurement uncertainty modeled as Gaussian white noise. The observation matrix is simply an identity matrix.

[0060] A problem exists when the left bound or right bound, or both, of a potential threat are outside of the camera's field of view. In such cases the potential threat bounds can not be directly measured from the stereo depth map. In such situations a very large variance is assign for the observation noise to reflect the uncertainty in the measurement. Experiments show that Kalman filtering propagates the uncertainty quite well. Kalman filtering is particularly helpful when a potential threat is very close and fully occupies the field of view. If the observations of bound positions and velocity become highly uncertain the collision detection system 102 relies on a previously estimated system model.

[0061] While the foregoing is directed to embodiments of the present invention, other and further embodiments of the invention may be devised without departing from the basic scope thereof, and the scope thereof is determined by the claims that follow.